不同標準設定方法之比較研究

謝名娟 / 謝進昌

不同標準設定方法之比較研究

謝名娟
國家教育研究院助理研究員

謝進昌
國家教育研究院助理研究員

摘要

在本篇研究中，比較書籤標定法與Yes/No Angoff標準設定方法在設定切斷分數上的差異，並使用TASA 2009年英文科來進行研究。研究結果顯示，兩種標準設定方法所做出的決斷分數略有不同，而成員們覺得Yes/NoAngoff法執行起來較為簡單，而最終決定的切斷分數也較符合預期，書籤標定法對於成員而言，需要克服的技術層面較多，尤其是在面臨題本難度的順序，和心中期望的難度順序不一致時，往往會造成他們很難找到確切的書籤位置，但是，書籤標定法所需的執行時間較短，做出的切斷分數結果，反應在2009年的實徵數據上，發現中間部分的學生比例明顯低於基礎以下與進階的學生，許多成員覺得做出切斷分數呈現M型分布，較能反應台灣學生學習成就的真實情形。

關鍵字：Yes/No Angoff方法、書籤標定法、標準設定

The Comparison of Different Procedures in Standard Setting.

Ming-Chuan Hsieh
Assistant Research Fellow, Research Center for Testing and Assessment, National Academy for Educational Research
mhsieh@mail.naer.edu.tw

Jin-Chang Hsieh
Assistant Research Fellow, Research Center for Testing and Assessment, National Academy for Educational Research

Abstract

This study compared two standard setting methods- bookmark and yes/no Angoff method, using 2009 TASA 2009 English as research subject. It is found that the resulting cutoff points from these two methods are somewhat different. Judges regarded that yes/no Angoff method is easier to implement and the resulting cutoff points are more close to their expectation. Comparatively, there are more difficulties need to be solved for the bookmark procedure, especially when the item difficulty order does not follow judges' expectation. When this kind of situation happened, it is very hard for the judges to place the bookmarks. However, the needy time for implementing the bookmark procedure is much shorter. The resulting cutoff point from bookmark method shows that most Taiwanese students centered at the below proficient or advance level, and not many in the middle level. Many judges regarded this kind of M shape distribution of student performance more closely reflected the reality of education condition in Taiwan.

keywords: yes/no Angoff, bookmark, standard setting

壹、研究背景及目的

臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement，簡稱TASA）的建置目的，為了解國小四年級、六年級、國中二年級、高中、高職二年級學生之學習成就表現，並探討學生學習成就上之表現差異與學習上變遷之趨勢，進而檢視當前國家教育體制與政策實施之成效。在此目的下，其中最重要的工作項目之一，即是評定或描述學生目前的學力表現，是否達到政府決策者所預期的程度或標準。所謂標準設定，其意涵在於測驗研發者或使用者根據測驗結果，將不同學生的表現加以分類，並確保某些符合通過資格的應試者能通過檢定或達到某一成就水準。

本研究針對2009年臺灣學生學習成就評量資料庫國民小學英文科的學習成就標準設定，並設定基礎、精熟、進階三個水平，為了瞭解標準設定的結果是否會受到不同設定方法的影響，本研究在TASA的英文科的標準設定採用兩種設定方法-書籤標定法與Yes/No Angoff標定法，來比較不同設定方法間所產生之決斷分數的差異。

貳、研究方法

本研究針對2009年臺灣學生學習成就評量資料庫國民小學英語文，來進行學習成就標準設定，並設定基礎、精熟、進階三個水平，為了瞭解標準設定的結果是否會受到不同設定方法的影響，本研究採用兩種設定方法，來比較設定方法間所產生之決斷分數的差異。

會議的第一天，進行的方法為Yes/No Angoff 方法，在這個方法中，成員們必須針對題庫中的每一個題目，來判斷是哪一種水平的邊緣受試者（borderline students，即為此水平中，最低成就的學生）才能夠答對。如果覺得基礎水平邊緣受試者可以答對此題，則在此題的基礎水平空格上打勾，同樣的，也需考量是否精熟與進階的邊緣受試者是否可以答對此題。

會議的第二天的成員和第一天相同，而進行的方法為書籤標定法，在這個方法中，先把題庫裡所有的題目由簡單到難依序排列（ordered item booklets，簡稱OIB），每一位標準設定成員檢視完試題之後，則須在OIB中依序放置基礎、精熟、進階的書籤。而放置書籤的原則，則以各水平中，最低成就的學生，應具備的英文科知識為主。然後對照OIB中的各個試題內容，找出哪一題為該水平的邊緣受試者，可以有67%的機率能正確作答的題目，並把書籤放在那個試題位置。

在這兩種方法中，每一輪判別完，研究者會提供各個成員所判別的決斷分數位置，以供全體成員討論。這樣的過程會重複三次，而成員們也可以在每一次過程中，修正自己的決定。第三輪之後，依據成員們在每個水平中所判定分數的平均數，計算最終的決斷分數。

兩種方法所判別的決斷分數，將會進行比較。此外，每一輪標準設定後，會對成員們進行問卷調查，以瞭解哪一種方法最符合成員們心目中理想的決斷分數。以下，就這兩種標準設定方法，加以詳加描述：

一、Yes/No Angoff 方法

Angoff方法為William Angoff（1971）所提出，Angoff方法使用上很簡單易懂，而且能夠輕易為不同形式的題型設定決斷分數。在最原始的Angoff方法中，標準設定成員，必須對OIB 中的每一個題目進行判別，並決定邊緣受試者（minimally competent examinee），有多高的機率，可以達對這個題目。然而，若是使用原先的Angoff 方法，除了要對每一題進行判斷之外，成員們還須思索每一題的答對機率，當題本內的題目很多時，這種方式就變得較不合適。

Angoff方法有許多修定的版本，其中一個廣泛使用的版本為Impara 與 Plake（1997）所設計的Yes/No Angoff 方法。這個方法，和原先設計雷同，必需要對題本中所有的題目進行判斷，但不同的點是，此方法不用寫出邊緣受試者答對題目的機率，而是直接寫下邊緣受試者是否能夠答對此題。如果邊緣受試者可以答對此題，則在這個題目的表格上寫下「是」，如果不能答對，則寫下「否」。這種較為直觀的判斷，減少了原先Angoff方法的執行難度。

Yes/NoAngoff 標準設定法的實際操作概念及流程，大致是如下所述。首先，研究者會事先提供每位標準設定成員一本英語文試題卷，而每頁的試題內容大致如下表1的範例所示，包含有題目內容、選項、答案及評測項目等。

表1 英語文某一範例試題格式

科別	年級	編號	答案
英語文	六	1	2
題目	【聽到】：book, book, book		
	【題目】：① good ② book ③ put		
能力指標	1-1-3能聽辨課堂中所習得的詞彙。		

註：本題目為經修訂之範例試題，非正式試題。

而後針對題本中的每一個試題，判斷是否基礎水平的邊緣學生，可以答對該題。若是不能，則判斷是否精熟水平的邊緣學生，可以答對該題，若是不能，則判斷是否進階水平的學生，是否可以答對該題。若是成員覺得此題很難，連進階水平的邊緣學生均無法答對，則可判定該題為超出進階水平的能力的題目。

每一位標準設定成員，逐題判斷之後，研究者將每一位成員的填答數據，輸入電腦軟體BILOG-MG中，並依據試題在題庫中的a, b, c 參數，固定試題參數來計算每個成員在基礎、精熟、進階三個水平上所估的能力值。將BILOG-MG估算出每一位成員在各個水平上所估的能力值加以平均，即可算出各水平的決斷分數。

第二輪執行的任務大致是重覆第一輪的動作，但差別在於研究者會提供一些回饋訊息，以作為成員參考，包含第一輪標準設定後，其它成員（與自己）所對各題的判斷、決斷分數。成員即依據回饋訊息，分成小組來討論第一輪所設定通過分數的適切性與聆聽其它成員發表自己對題目的判別依據，進行第二輪的設定，並再次對各題進行判定。第二輪標準設定結束後，所提供的回饋訊息為其它成員（與自己）所對各題的判斷、決斷分數，與依此決斷分數，使用2009的學生真實成績，來

計算各水平的通過人數百分比。成員再次依據這些回饋訊息，重新進行每一個題目的判定。

第三輪執行的任務亦是雷同於前兩輪，但差別在於此輪任務是由成員獨自完成，不能相互討論，最後，研究者即是根據標準設定成員於第三輪所判定的成果，經換算後，以得基礎、精熟與進階水平的正式通過分數。

二、書籤標定法

書籤標定法（bookmark）（Lewis, Mitzel, & Green, 1996）的設計不僅能避免原始Angoff法逐題檢視、評定之疲勞、繁瑣干擾，同時，也能輕易的加入選擇題型與建構反應試題來進行標準設定，而在執行這個方法時，研究者會事先提供每位標準設定成員一本已經藉由IRT預先計算出的難度值，並加以由簡單至困難排序的英語文試題卷（OIB）。由於TASA的所有技術報告均以試題反應理論三參數進行分析，為了使標準設定的結果能應用於TASA，對於難度值的計算，是採用三參數試題反應模式來進行估計。

標準設定成員逐一檢視OIB中所有試題後，將基礎、精熟、進階三個書籤逐一放置於某一試題上，即完成三個水平通過分數的設定，然而，放置書籤位置時，須依照標準表現描述對於各水平學生的描述進行推理和想像，並推測該水平中，程度最差的那一位學生，具有67%的機率能該答對該題，並把書籤放置在那個試題的位置上，記錄在記錄表中。例如：若將基礎書籤放置在第16題、精熟書籤是放置在第40題、進階書籤是放置在第70題。即可將所有學生區分出基礎以下、基礎、精熟與進階等四個能力區塊。接續，研究者會將各水平書籤所放置對應的試題，搭配該題的已知的試題難度、鑑別度及猜測度參數，於反應機率（response probability）0.67（Huynh, 2006）下，進行能力值的轉換，如此，即是該水平的通過能力。和Yes/No Angoff方法相同，成員總共要進行三輪來放置書籤，第一輪放置後，研究者會提供一些回饋訊息，以作為成員參考，並進行討論，以了解設定通過分數的適切性，而後，成員再依據前述的原則，重新放置第二輪的基礎、精熟、進階三個書籤的位置。第二輪後提供回饋訊息，進行成員討論後與第三輪的標準設定，但提供的回饋訊息，可略有不同，以了解不同回饋訊息，對於成員設定分數的影響。

最後，研究者即是根據標準設定成員於第三輪所放置於基礎、精熟、進階三個書籤的位置，經換算後，以得基礎、精熟與進階水平的正式通過分數。

三、標準設定成員

本研究發出標準設定成員徵求訊息後，經為期二個多月的成員甄選，正式參與TASA英語文小六標準設定成員總人數為32名，而成員是分佈在北部（20名，62.5%）、中部（7名，21.9%）、南部（2名，6.3%）、東部（3名，9.4%），其中教師占有24名（75%），其次為行政人員4名（12.5%）、學者4名（12.5%），而性別的分佈是男性占有6人（18.8%）、女性占26人（81.2%）的比率。最後，小六成員總教

學年資或行政年資，最低是2年、最高是31年，平均年資是10.2年。整體而言，TASA英語文小六標準設定成員大致能含括北中南東四個區域的人員、同時能兼顧教師、行政人員及學者的代表組成，最後，教學或行政年資具高水平，平均皆超過10年。

四、會議流程

這32名選定的標準設定成員，應邀參加兩天的標準設定會議。 在會議進行前的一個禮拜，研究小組先寄送會議的前導資料，包括標準表現描述、評量架構、會議簡介、與會議流程說明。會議第一天，所邀請之32位成員全數出席，先由研究者簡要說明會議的目的、流程之後，並請所有成員，就標準表現描述內的細項內容，逐一檢視，並加以討論，並請TASA英文科召集成員協助釐清成員們的疑問。之後，成員檢視題本，並進行第一輪Yes/No Angoff方法的標準設定，第一輪結束後，研究小組進行統計分析，並繪製每一位成員給定分數的散佈圖、各試題傳統難度P值，與每一題，成員給定基礎、精熟、進階的比例。成員們就回饋訊息的內容，逐題進行討論，原先設計30分鐘進行小組討論，30分鐘全體討論，由於題目眾多，成員無法在30分鐘內完成小組討論，因此要求取消全體討論時間，以求能盡量能與小組成員討論題本內的題目。小組討論後，成員們修正彼此的看法與意見並填寫執行方法的評估問卷，內容包括對於標準設定方法的理解，或是對於會議流程進行的建議等。而後進行第二輪的判定，判定後則同樣提供成員相同的回饋訊息，並再次進行小組討論與填寫回饋訊息的評估問卷，問卷內容為對於回饋訊息的理解度等。最後進行第三輪的標準設定。 設定的結果公布給成員之後，進行成果問卷的填寫，問卷內容包括對於成員們對於自己所設定的分數信心強度、覺得最終結果是否合理等，會議結束後研究者將所有題本、回饋訊息回收。

會議第二天為同一批成員，除了有一位成員因病缺席之外，其它31位委員均參與第二天的會議。第二天所執行的標準設定方法為書籤標定法。流程與第一天雷同，第二天採用同樣的會議資料來進行書籤標定法的標準設定。但會議前與成員們強調題本是依據三參數試題反應理論中的難度，由易至難進行排序，由於三參數的難度估計，易受鑑別度和猜測度的影響，因此在題本難度的排序上，可能和成員們的心中的期望的難度排序不完全吻合。

第二會議流程，和第一天相同。但在回饋訊息的提供上略有不同，第一輪結束後，提供的回饋訊息為每一位成員所給定書籤位置的散布圖，而後進行30分鐘小組討論與30分鐘全體成員討論。第二輪結束後，除了提供每位成員給定書籤位置的散布圖之外，應某部份成員要求，呈現依據成員所給訂的切斷分數，在2009 年TASA的實徵數據下，有多少百分比會落在基礎以下、基礎、精熟和進階四個水平，研究者解釋完回饋訊息之後，則進行30分鐘小組討論與30分鐘全體討論。第二天會議成員評估問卷和第一天雷同，問卷中除了詢問關於書籤標定法所設出的結果評估之外，亦有幾題是詢問成員對於兩種標準設定法執行後的感想。

參、研究結果

一、切斷分數

表2與表3呈現在每一輪各水平的切斷能力值。對於書籤標定法而言，精熟水平在三輪中的書籤標定位置產生的波動較大，由第一輪的-0.34，第二輪的-0.43，到第三輪的-0.08，相對而言，對於基礎水平和進階水平而言，相對變動的幅度較小。對於Yes/No Angoff法而言，進階水平的切斷分數則於第二輪到第三輪的變動的幅度較大。相對而言，基礎與精熟的切斷分數在每一輪的變動幅度較為穩定。此外，書籤標定法在判定每一水平的切斷分數上，成員們評定分數的標準差較低，代表成員們在執行書籤標定法時較能達到共識，但唯一的例外為精熟水平的標準差，達到0.5左右，代表成員們在經過第二輪的討論之後，對於精熟的書籤標定位置有較大的分歧。

最終的切斷分數是以第三輪的結果來判定，就第三輪的切斷分數來看，Yes/No Angoff法在基礎與精熟水平的切斷分數略低於書籤標定法，但是進階水平的分數卻略高於書籤標定法有1個logit 之多。

表2 書籤標定法結果

輪	決斷分數（標準差）		
	基礎	精熟	進階
1	-0.90 (0.16)	-0.34(0.28)	0.21(0.22)
2	-0.94 (0.11)	-0.43(0.23)	0.30(0.17)
3	-0.96 (0.06)	-0.08(0.50)	0.22(0.10)

表3 Yes/No Angoff法結果

輪	決斷分數（標準差）		
	基礎	精熟	進階
1	-1.53 (0.42)	-0.41(0.28)	1.59(0.38)
2	-1.44 (0.22)	-0.35(0.27)	1.69(0.27)
3	-1.51 (0.28)	-0.44(0.00)	0.44(0.36)

表4為依據每一輪在各水平的切斷分數，並將此切斷分數放入2009年TASA的實徵數據中進行運算各階層的百分比人數。這兩個表中可以看出對於書籤標定法而言，精熟水平在每一輪中的變化較大，由第一輪的18%，第二輪的25%，到第三輪的11%。而對於Yes/No Angoff法而言，則在進階水平的變化較大。之所以產生變化的原因，在於進行Yes/No Angoff時，第一次的回饋訊息，單單只提供每個成員所給定判斷值得相對位置散布圖與每一題的傳統難度P值，而在第二次回饋訊息時，除了第一輪給定的回饋訊息之外，還呈現了各階層的百分比人數，許多成員在看到進階水平的人數過少，覺得如此的切斷分數會造成社會觀感不佳的問題，因此普遍降低原先所設定的進階水平切斷分數。

進行完兩天的標準設定之後，研究者要求成員寫下他們心目中理想的各階層人數的分配比率，所得的各階層分配如表五，由這此分配而言，較為接近Yes/No Angoff方法所設出的結果。 然而，在進行訪談時，多位成員表示，書籤標定法所設

出的結果反應出教學現場的真實情況，即為一個M型的學習生態，大多的學生集中在非常優秀或是極為落後的兩群，而中間程度的學生比率相當少，因此，成員們覺得雖然Yes/No Angoff法所設出的成果，較貼近表現標準描述與社會的期待，但是書籤標定法所呈現的結果，卻能反應教學現場的真實情況。

表4 各階層的百分比人數-書籤標定法、Yes/No Angoff法與成員心目中的理想比率。

水平	第一輪		第二輪		第三輪		成員的理想比率
	bookmark	Angoff	bookmark	Angoff	bookmark	Angoff	
基礎以下	20	8	19	9	19	8	13
基礎	16	26	14	27	25	25	24
精熟	18	65	25	64	11	34	34
進階	46	1	42	0	45	33	29

在標準設定的會議進行中，成員們就整體的會議說明、方法執行、回饋訊息的提供等各項層面對兩個標準設定方法進行評估，評分採李克氏量表， 1代表態度為正向的程度為最低（如 非常不同意、非常不清楚、非常沒信心等）、5為態度為正向的程度最高。而 1到 5中間的數字，則代表程度上的差別。例如1代表非常不同意、2代表不同意、3代表沒意見、4代表同意、5代表非常同意等。

以下就這幾個層面，來進行說明：

（一）會議說明

成員們認為收到前導資料能夠有助於了解會議中，自己所應扮演的角色，對於TASA的測驗目的與標準方法的執行，都呈現極高的評價與理解。然而，成員們對於Yes/No Angoff方法，給予的評價高於書籤標定法。例如在「我已經了解Yes/NoAngoff方法，並可以使用這個方法，進行試題的判別與歸類」，成員們給的平均分數為4.06，但對於「我已經了解Yes/No書籤標定法，並可以使用這個方法，進行試題的判別與歸類」，則給的分數為3.84。 然而、對於會議流程的進行來說，書籤標定法的分數高於Yes/No Angoff方法，例如「我瞭解會議接續的標準設定流程」，Yes/No Angoff方法為4.13，書籤標定法則為4.26。這是因為第一天執行的方法為Yes/No Angoff方法，而第二天執行的方法為書籤標定法，因兩天的流程都相同，成員對於自己要進行的任務，在第二天也有較清楚的認識。

（二）方法執行

兩個方法執行上來說，成員們覺得Yes/No Angoff方法，執行上比較容易，也較能夠和PLD作連結。例如在「基礎的表現標準描述（PLD）有助於我判別邊緣基礎受試者可以答對的題目」，成員給Yes/No Angoff方法的平均分數為3.78，而在「基礎的表現標準描述（PLD）有助於我置放介於邊緣低於基礎/基礎的書籤」，的平均分數為3.43，相似的，精熟與進階也呈現類似的情況。此外，成員們覺得Angoff方法比書籤標定法，更容易理解與應用自己的教學經驗。然而，由於Angoff方法必須逐題判斷，許多成員覺得討論的時間不夠充裕。相對而言，書籤標定法的給予成員們的討論時間就較為充裕。

（三）回饋訊息與討論

對於回饋訊息而言，就各成員給定分數的散佈圖、試題難度、各水平百分比而言，成員們都很了解個訊息的意涵，然而，成員對於Yes/NoAngoff 的回饋訊息理解程度，還是略高於書籤標定法，例如，成員對於其它成員（與自己）判定決斷分數的散布圖說明，Yes/No Angoff方法的清楚度，達到4.06，而對於書籤標定法，則為3.84。但成員們覺得書籤標定法的結果較容易討論，因為在只需要討論三個書籤的放置位置，但Yes/No Angoff法卻必須要逐題討論，很多成員都反應討論時間不夠。

（四）兩種設定方法的比較

成員普遍認為Yes/No Angoff標準設定法，是比較簡單、易懂的執行方法，在「哪一種標準設定方法較為簡單易懂」 與「哪一種標準設定方法較為易於執行」上，有27位的成員選擇Yes/No Angoff法， 3位選擇書籤標定法，而1位沒有意見。此外，有30位成員覺得Yes/No Angoff法，較能與標準表現描述作結合，最後，對於標準設定法所設出的結果，26位成員覺得Yes/No Angoff法所做出的結果較符合自己心中的預期，只有4位覺得書籤標定法所做出的結果比Yes/No Angoff 法較符合心中預期。

詳細的問卷內容的比較請詳見表5。

表5 Yes/NoAngoff法與書籤標定法評估問卷比較

題目		Angoff	書籤標定
會議說明	我認為先前收到的前導資料能充分幫助我瞭解本次會議應扮演的角色	4.34	4.19
	我瞭解TASA的測驗目的與施測對象	4.35	X
	我瞭解本次TASA標準設定會議的目的	4.41	4.16
	本次會議對於表現標準描述（PLD）的說明及其範例的陳述，我認為	4.00	X
	本次會議對於如何執行Yes/No Angoff／書籤標定法的說明，我認為	4.06	3.94
	我已經了解 A Yes/No Angoff方法，並可以使用這個方法，進行試題的判別與歸類。 B書籤標定法，並可以使用這個方法，進行標準設定的工作	4.03	3.84
	我瞭解邊緣受試者的涵義	3.97	4.13
	我瞭解會議接續的標準設定流程	4.13	4.26

方法執行方法執行	本次會議的解說與導引時間分配，我認為	3.16	3.06
	本次會議提供歸類試題／置放書籤的時間分配，我認為	2.94	3.23
	基礎的表現標準描述（PLD）有助於我 A判別邊緣基礎受試者可以答對的題目 B置放介於邊緣低於基礎/基礎的書籤	3.78	3.43
	精熟的表現標準描述（PLD）有助於我 A判別邊緣精熟受試者可以答對的題目 B置放介於邊緣基礎/精熟的書籤	3.56	3.45
	進階的表現標準描述（PLD）有助於我 A判別邊緣精熟/進階受試者可以答對的題目 B置放介於邊緣精熟/進階的書籤	3.78	3.47
	研究者所提供依難度排序試題本（OIB）符合我所知覺試題間的相對難度	X	2.68
	我對採用67%的正確作答標準，去界定書籤的位置，感到合適	X	3.39
	我先前的教學經驗，有助於我瞭解 A進行試題的判別與歸類 B該如何置放各階層的書籤位置	4.28	3.81
回饋說明與討論	對於其它成員（與自己）判定決斷分數的散布圖說明，我認為	4.06	3.84
	我瞭解其它成員（與自己）判定決斷分數的相對位置	4.13	4.00
	對於基礎、精熟、進階等水平通過人數百分比的說明，我認為	4.03	4.00
	我瞭解基礎、精熟、進階等水平通過人數百分比的意涵	4.25	4.00
	對於試題通過率的說明，我認為	4.31	X
	我瞭解試題通過率的意涵	4.45	X
	基礎、精熟、進階等水平通過人數百分比會影響我 A對試題的歸類 B放置書籤的位置	3.81	3.71
	其它成員所判定之決斷分數會影響我 A對試題的歸類 B放置書籤的位置	3.44	3.42
	試題通過率會影響我對試題的歸類	3.84	X
	試題品質的好壞（如誘答選項的設計，或是題目敘述不清）會影響我 A判別各水平受試者可以答對的題目 B放置書籤的位置	4.50	4.61
	透過小組的討論，我充分瞭解其他成員的想法	4.25	4.26
	透過與小組成員的討論，有助於我 A對試題的歸類 B放置書籤的位置	4.19	3.97
	本次會議提供小組討論標準設定結果適切性的時間分配，我認為 1.太短 2.略短 3剛剛好 4略長 5太長	2.59	3.32

結果評估與方法比較時間分配設定方法	我相信自己對每一個試題／水平的判別，是與表現標準描述（PLD）一致	3.78	3.23
	我對於自己所設置的決斷分數，深具信心	3.97	3.81
	我對於最後的決斷分數，感到	2.71	3.14
	本次會議各階段任務的執行時間，我認為 1.太短 2.略短 3剛剛好 4略長 5太長	3.03	3.28
	哪一種標準設定方法較為簡單易懂	27人	3人
	哪一種標準設定方法較為易於執行	27人	3人
	表現標準描述（PLD）對於執行哪一種標準設定方法較有幫助	30人	1人
	哪一種標準設定方法所設出的結果，較符合您的預期	26人	4人

註：X代表該天問卷中沒有這個題目。評分採李克氏量表，除非特別標明評分的尺度，否則1代表態度為正向的程度為最低（如非常不同意、非常不清楚、非常沒信心等）、5為態度為正向的程度最高。而1到5中間的數字，則代表程度上的差別。例如1代表非常不同意、2代表不同意、3代表沒意見、4代表同意、5代表非常同意等。

肆、結論與建議

在本篇研究中，比較書籤標定法與Yes/No Angoff標準設定方法在設定切斷分數上的差異，並使用TASA 2009年英文科來進行比較的研究。研究結果顯示，兩種標準設定方法所做出的決斷分數略有不同，而成員們覺得Yes/No Angoff方法對他們來說是執行起來較為簡單的方法，而最終決定的切斷分數也較符合預期，書籤標定法對於成員而言，需要克服的技術層面較多，尤其是在面臨題本難度的順序，和心中期望的難度順序不一致時，往往會造成他們很難找到確切的書籤位置，但是，書籤標定法所需的執行時間較短，對於成員的心理與體力負擔較輕，做出的切斷分數結果，反應在2009年的實徵數據上，發現中間部分的學生比例明顯低與基礎以下與進階的學生，許多成員覺得做出切斷分數呈現M型分布，較能反應台灣學生學習成就的真實情形。

此外，在標準設定中，標準表現描述（PLD）為成員們達成共識的主要依據，然而，就應用層面而言，成員們反應Yes/No Angoff方法較容易和PLD進行連結，因為Angoff方法為逐題判斷，成員們比較容易一題一題找出哪一題為基礎水平的邊緣學生可以達對的題目，哪一題為精熟水平的邊緣學生可以達對的題目。然而，對於書籤標定法就比較不容易了，因為要找出三個切斷點，基礎的切斷點代表基礎水平的邊緣學生，大概能答對所有的題目，可是對於OIB題本，是依造試題反應理論的難度來排的，並不是依據標準表現描述來排的，因此，成員反應雖然可能將切斷點到某一試題，但是並不代表此試題之前的所有題目，該水平的邊緣學生都可達對。例如，基礎的切斷點是放置在第20題，但是有可能第 5、 8題，是精熟學生才能答對的。因此，成員進行書籤標定法時，與到許多困難與衝突點，但透過討論後，的確能加速成員達成共識。

對於回饋訊息方面，成員們對於自己與其他成員的切斷分數散布圖感到十分重視，因為許多成員都不希望自己成為「異類」，因此，看到散布圖後，均會修正自己原先的判定，而逐漸和其他人判定的結果相同。但是，在兩天的標準設定會議中，卻看到明顯的城鄉差距。城市的學校老師，所設定的標準較高，而偏遠的學校老師，則設定的標準較低，即使設定時是依據PLD來進行推論，但是教師們還是會加入自己在教學現場經驗和看法，來進行標準設定。

就成員們評估的結果而言，Yes/No Angoff方法比較適合TASA英語科的標準設定，然而，進行此方法的前提為要有足夠的會議時間進行討論。就本研究原先設計，為每一輪成員進行標準設定的時間為50分鐘，討論為40分鐘，但是，由於Angoff的方法要逐題討論，因此，對於題本題目較多的測驗，是有執行上的困難的。例如TASA英文科的題本考題有一百多題，成員大多沒辦法在 40分鐘內完成討論，造成許多成員是邊進行標準設定，邊進行討論。此外，在成員的數據輸入和統計分析上也是一大挑戰，本研究總共有 32位成員，每一位成員進行103題的判斷，因此要進行三千多筆數據的輸入。本研究動員5-6位助理同時進行輸入與檢誤，花費約30-40分鐘的時間，因此，使用Yes/No Angoff方法雖然比較簡單，但是要考量到時間與人力的負荷。 相對而言，書籤標定法的討論就很簡單，成員們無須逐題討論，只需討論自己放置書籤的位置就可以了，同樣安排討論的時間為40分鐘，而成員們只需30分鐘左右就討論好了，而輸入更為簡單，只要為每一位成員輸入三筆數據，本研究的書籤標定法的成員為31人，總共需要輸入的數據不到 100筆，本研究只有動員兩位研究助理，花不到10分鐘的時間就完成輸入和檢誤。因此，就題目較多的OIB或是人力不足的研究小組，較適合使用書籤標定法。

隨著國際的趨勢，國內雖然漸漸知覺到長期學習成就評量資料庫建置的重要性，但對於標準設定的議題，瞭解可說是少之又少，而本研究的研究過程與結果，除了做為TASA實務的運用，也可作為其他國內建置大型學習成就評量資料庫標準設定之參考。

參考文獻

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (pp.508-600). Washington, DC: American Council on Education.

Huynh, H. (2006). A clarification on the response probability criterion RP67 for standard.set tings based on Bookmark and item mapping. *Educational Measurement: Issues and Practice, 25*(2), 19-20.

Impara, J. C., & Plake, B.S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement, 34*, 353-366.

Lewis, D. M., Mitzel, H.C., & Green, D. R. (1996). *Standard setting: A bookmark approach.* Paper presented at the Council of Chief State School Officers National Conference on Large Scale Assessment, Boulder, CO.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark method: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.